MODEL EXPLANATION DOCUMENT

Early Warning System for Student Burnout & Dropout Risk

AI-Powered Behavioural Analytics Framework

Problem Understanding

1. Introduction

Universities often identify academic underperformance only after significant decline in grades or attendance. By the time formal intervention occurs, student burnout and disengagement may already be severe.

Behavioural analytics enables early detection of subtle patterns that signal academic disengagement before visible failure occurs.

2. Problem Objective

The objective of this project is to develop an AI-powered behavioural analytics system that:

- Detects early behavioural signs of burnout
- Predicts student dropout risk
- Generates a risk score (0–100)
- Identifies key behavioural triggers
- Recommends appropriate intervention strategies

3. System Output

The system provides:

- Risk Score (0–100)
- Risk Category (Low / Medium / High)
- Top behavioural indicators influencing prediction
- Suggested intervention strategy
- Interactive visualization dashboard

Dataset Description & Assumptions

1. Dataset Type

Dataset Type: Synthetic

2. Justification for Synthetic Data

Real university LMS behavioural datasets are not publicly available due to:

- Student privacy regulations
- Institutional data protection policies
- Ethical and confidentiality constraints

Therefore, a realistic synthetic dataset was generated to simulate behavioural patterns observed in academic environments.

3. Dataset Size

- Total Records: 2000 students
- Behavioural Features: 7
- Target Variable: Burnout Risk Level

4. Feature Description

Feature	Description
LMS_Login_Frequency_Per_Week	Weekly LMS activity frequency
Assignment_Delay_Days	Delay in submission
Attendance_Rate_Percentage	Class attendance rate
Study_Consistency_Index	Behaviour regularity indicator
Feedback_Sentiment_Score	Emotional tone indicator
CAT_Marks_Trend	Academic performance trend
Late_Night_Activity_Ratio	Activity irregularity indicator

5. Data Generation Assumptions

- LMS activity follows a normal distribution
- Assignment delay follows a Poisson distribution
- Attendance is uniformly distributed between 60–100%
- Sentiment is uniformly distributed between -1 to 1
- Academic trend ranges between -15 to +15

Feature Engineering & Model Selection

1. Feature Engineering Logic

The following preprocessing steps were performed:

- Target variable encoding (Low=0, Medium=1, High=2)
- Behavioural threshold rule creation for label assignment
- Probability-to-risk-score transformation
- Feature importance extraction from trained model

2. Risk Score Calculation

Risk Score = Maximum predicted class probability × 100

This provides an interpretable continuous score from 0 to 100.

3. Model Selection

Model Used: Random Forest Classifier

4. Reason for Model Selection

Random Forest was selected because:

- It captures nonlinear behavioural interactions
- It is robust to noise
- It performs well on structured tabular data
- It provides feature importance for explainability
- It reduces overfitting compared to single decision trees

Alternative models such as Logistic Regression were not selected due to their limited ability to model complex behavioural interactions.

Model Performance & Evaluation

1. Train-Test Split

- 80% Training
- 20% Testing

2. Evaluation Metrics

- Accuracy: 98.75%
- Precision
- Recall
- F1-Score
- Confusion Matrix

3. Interpretation

The model demonstrates strong predictive capability across all classes.

- Medium risk students were identified with near-perfect precision.
- High-risk students were correctly identified with strong recall.
- Low-risk misclassifications were minimal.

High accuracy is expected due to synthetic rule-based label generation, validating the relationship between behavioural signals and burnout risk.

Behavioural Insights & Practical Feasibility

1. Behavioural Insights Derived

Top 3 Behavioural Triggers Identified:

1. CAT Marks Trend
2. Attendance Rate
3. Feedback Sentiment Score

Key Findings:

- Declining academic performance is the strongest burnout indicator.
- Attendance reduction strongly correlates with disengagement.
- Negative emotional sentiment increases dropout probability.

2. Intervention Strategy Logic

The system provides automated recommendations:

High Risk:

- Immediate counselling
- Faculty advisor notification
- Weekly monitoring

Medium Risk:

- Academic reminders
- Peer mentoring

Low Risk:

- No intervention required

3. Practical Feasibility

This system can be integrated into:

- University LMS systems
- Academic monitoring platforms
- Institutional analytics dashboards
- Student counselling frameworks

It enables proactive intervention rather than reactive correction.

4. Conclusion

This project demonstrates that behavioural analytics combined with machine learning can proactively identify student burnout risk.

By converting behavioural signals into predictive intelligence, universities can:

- Reduce dropout rates
- Improve academic engagement
- Support student wellbeing
- Enable data-driven decision making

The system transforms reactive academic monitoring into an early warning behavioural intelligence framework.